09813955.032201

CLAIMS

1.A method of recognizing and indexing documents in a system having a scanner (30) connected to a computer, the method comprising:

scanning the documents;

using a pointing device or member of the computer to designate an arbitrary point P in at least one box of the documents;

recognizing by OCR the characters in said box; and

storing the recognized characters in a first database connected to the computer to enable documents scanned in this way to be indexed.

2. The method according to claim 1, wherein said designation step comprises searching for and identifying the box of the document which contains said point P designated by the user.

3. The method according to claim 2, wherein said step of searching for and identifying said box is performed by applying a shape search algorithm over a determined search zone surrounding said point P as previously designated by the user.

4. The method according to claim 3, wherein said shape search algorithm is a projection algorithm which counts the number of pixels present in each vertical or horizontal line of said determined search zone and which, on the basis of these count numbers, finds the horizontal and vertical lines present in said search zone by examining the peaks in the X and Y projection profiles.

5. The method according to claim 3, wherein said shape search algorithm is an algorithm based on the Hough transform.

6. The method according to claim 1, wherein said OCR step is preceded by a step in which the user defines the type of character to be recognized in said box of the document.

7. The method according to claim 1, wherein said scanning step is performed initially for a set of documents to be processed, with said steps of identifying the

box and performing OCR being performed subsequently in succession for each of the documents.

8. The method according to claim 1, wherein said scanning step is initially performed for a first document, with said steps of identifying the box and performing OCR subsequently being performed on that document so as to define a sequence, with said sequence then being repeated in succession for each of the documents to be processed.

9. The method according to claim 1, wherein said documents to be recognized and indexed are a set of technical drawings of the same or different types.

10. The method according to claim 1, wherein said documents to be recognized and indexed are a set of forms, of the same or different types.

11. An apparatus for recognizing and indexing documents, the apparatus comprising:

a scanner for scanning a document and delivering an image of the document;

a computer connected to the scanner to receive said scan image;

a first database connected to said computer for storing said scanned image; and

first software for using a pointing device or member of the computer to designate an arbitrary point P in at least one box of the image, for searching for and identifying the box containing said point P designated by the user, for recognizing by OCR the characters in said box, and for storing the recognized characters so as to enable images scanned in this way to be indexed.

12. The apparatus according to claim 11, further comprising a second database connected to the computer to store characterization data such that the box subsequently can be identified automatically by said software without any point P within said box being designated.

13. The apparatus according to claim 11, wherein further comprising second software for defining the type of data to be recognized in said document box.

09813955.032201

14. The apparatus according to claim 12, wherein the first and second databases are integrated in the memory of the computer.

15. The apparatus according to claim 11, wherein said pointing member is the keyboard of the computer or a finger of the user.

16. A computer-readable medium having embodied thereon software to be processed by a computer connected so as to receive a scanned image, the software being operable to cause said computer to perform the functions of the first software of claim 11.

09813955.032201